Data Analytics For Earth Science

Pooja Sharma
Department- Data Science
Atlantic Technological University
Letterkenny, Ireland
L00171181@atu.ie

***Abstract*—A new discipline called "Data Science" has emerged as a result of the development of sophisticated e-infrastructures, which have made enormous storage and processing capacities possible. Big Data Analytics and DS techniques are anticipated to facilitate the processing of the numerous Earth Observation datasets that are now being gathered and created through observations and simulations, which will likely be advantageous to Earth and Environmental sciences. The special characteristics of Earth Science data and applications, however, are the significance of the geographical informations, the wide range of data types and formats, and the complexity of processing. As a result, Big Earth Data Analytics(BEDA) calls for approaches and instruments that have been specially developed. The Earth-Server Big Earth Data Analytics engine, which is based on high-performance array database technology and the acceptance and advancement of service interface standards, offers a solution for coverage-type datasets. With EarthServer, you can access information via mobile phones and visualize data via a variety of query languages and scalability options, all powered by a set of standards developed by global scientific collaborations. An illustration and explanation for this claim is the proliferation of lighthouse applications in the fields of marine geology, atmospheric, extraterrestrial, and natural climate science.**

Keywords: DS; BDA; EarthServer; BEDA Analytical Methods

1. INTRODUCTION

As communication and digital storage technologies have evolved, a great deal of information has been collected, leading to the need for more efficient ways to maintain, store, and process information. In this setting, the phrase "big data analytics" spread like wildfire. It was said that the volume, quickness, and diversity of this data need outstanding managerial skills. A new approach to science, called e-Science, is tackling big data challenges in several scientific disciplines. The disciplinary domains pushing the e-Science approach the most and perhaps benefiting from it include the earth sciences. Because of the significant storage and computational challenges that Earth Sciences bring, such as

1. They are used for a range of things, including the interdisciplinary study of the Earth as a system and disciplinary sciences (including climate, oceanography, and geology).
2. In order to quickly handle complicated situations (such as those involving climate change), they employ sophisticated scientific models and simulations (Big Velocity).

2. Defining Earth Science Data Analytics

DS research and create tools for processing, storing, and displaying data. They are executing data analytics when they put their knowledge to use on specific challenges, using tools and methodologies to co-analyze heterogeneous data. Researchers, developers, and data analytics practitioners all need the same skill sets as data scientists. People may use ESDA(Earth Science Data Analytics) to find patterns, correlations, and other data to learn more about our planet.

It consists of:

1. Data Preparation: Preparing Heterogenous data
2. Data Reduction: Correcting, Ordering, and Simplifying data
3.Data Analysis: Applying techniques and Methods for results.

3. Big Earth Data Analytical Methods

After preprocessing, data analytics' primary goal is to unearth hidden patterns of heterogeneous data to support Earth Scientific Research. The associated techniques may be divided into machine learning, deep learning, statistics, and model simulation and prediction.

Model Simulation and prediction: General circulation models (GCMs) provide predictions of both the planet's atmosphere and global ocean circulation, which aid in climatic research and climate change projection.

Traditional statistical Methods: To help us better understand the climate and estimate climate change, general circulation models anticipate both the planet's atmosphere and the global ocean circulation.

Machine Learning Methods: To infer intractable problems and uncover previously overlooked patterns, machine learning approaches build models based on qualities and attributes gathered from empirical data.

Deep Learning Methods: DL has achieved exceptional performance in computer vision(CV), natural language processing(NLP), and recommendation systems(RS) as a

result of its more sophisticated expression and optimization capabilities.

Time Series Analysis: Time series analysis is an additional method or statistical tool for data analysis.Time Series data

- Cross-Sectional Data
- Pooled Data

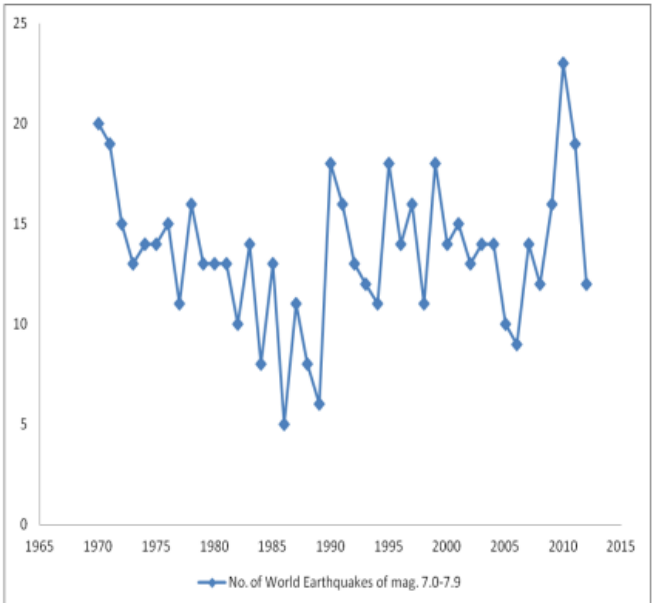

Figure 1: Analysis of Time Series

4 Expanding Earth Science Data Systems Using Cloud Computing

DS Systems enable the effective, organized, and large-scale processing, testing, and analysis of instrument observational data. Remote sensing satellite views of the Earth are sent to a variety of data systems on the planet for processing, storage, and analysis. But as instruments, computer systems, scientific requirements, and algorithm complexity has increased, so has the amount of data that has to be processed.

Traditional science data systems and data processing techniques are therefore no longer relevant. The increased data generated by these new forcing functions calls for innovative solutions, such as the consolidation of all end-to-end data systems into a single cloud computing region.

These increased needs and financial restrictions are compelling the organization to complete activities for migrating science data systems and other important data systems to the cloud.

5. Data Discovery and Analytics:

Traditional data discovery relies on open-source technology before initiating any data analytic activity. The metadata for this data is usually saved in a full-text search engine that functions to a Google search engine.

For large data analytics, distributed computing technologies are widely used in a variety of current systems. Two common open-source distributed options for large data analytics are Apache Spark and Hadoop MapReduce.

Simple mathematical procedures to faster and vector operations, visualization and exploration, and ML are among the most complicated analytical functions for earth data.

6. Tools & IDE for Analysis Earth Analytics in Computer:

IDE- Anaconda
Python, R
GeoMAC_DL,
OpenTopoDL
Earth Analytics Python Conda Environment
Python Library: EarthPy, Matplotcheck
R Package: qtoolkit-smapr(QS)
Cloud Computing

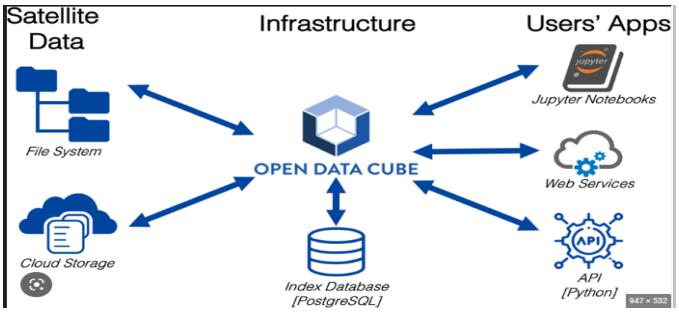

Figure 2: Scientific Data Structure

7. The Architecture & Design

Each Earth data system has its unique system strategy and architectural design.

These three domains—technological, scientific, and application—are illustrated by this architecture.

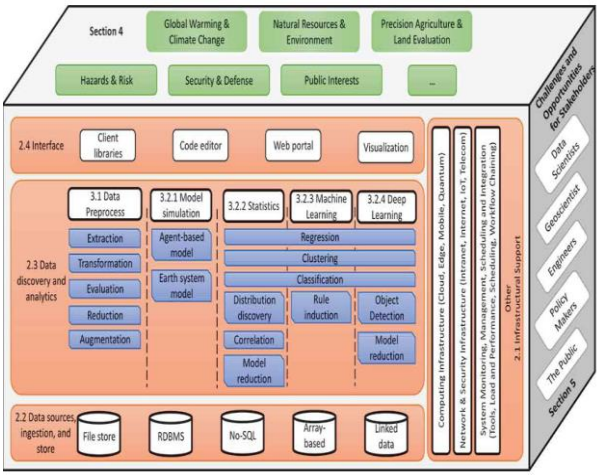

Figure 3.BDA Architecture's

8 **Support for infrastructure:**

Most BED(Big Earth Data) analysis systems have already been moved to a cloud computing environment or are in the process of doing so. Recent computer innovations have also moved some computing functions to the periphery of the infrastructure, such as mobile devices and smart objects at the edge of the IOT.

Computer infrastructure is required to support big data analytics, but tracking, measuring, accounting, and integrating facilities, as well as the network and security infrastructure, enable computing and analytics to be managed in a fluid, dynamic, secure, and user-friendly manner.

8.1 **Sources, intake, and storage of data:**

The data stores, which manage access to archived Earth data, is another crucial component of the system design. Data about the Earth are often kept in file systems, relational databases, or No-SQL databases, depending on their nature and intended use.
Data storage may normally be accessed in one of two ways:

- batch import;
- real-time import;

For instance, an Earth data search engine continuously gather data from the Internet and saves it at a predetermined place.

9. **With regard to BEDA, the EarthServer Method**

A network of open large-scale Earth data providers called EarthServer includes research institutions, supercomputing facilities, businesses, and government organizations.

As a single source of Earth data and services, EarthServer provides a single common information space for multidimensional spatiotemporal Earth data, to which all members of EarthServer collectively contribute.
The EarthServer project, funded by EU-FP7 infrastructures, assembles Eleven partners from the computer and earth sciences to create Big Earth Data Analytics(BEDA).

Earth science is described as the digital representation of data such as 1-D sensor time-series, 2-D remote sensing imaging, 3-D x/y/z geology data, and 4-D x/y/z/t atmosphere and ocean data in Big Data Analytics.

The EarthServer approach implements semantics-based dynamic query piece distribution in accordance with network optimization and additional criteria.
Any size, any type multidimensional raster data may be effectively stored and retrieved using the rasdaman Array DBMS, which is a component of the EarthServer platform.

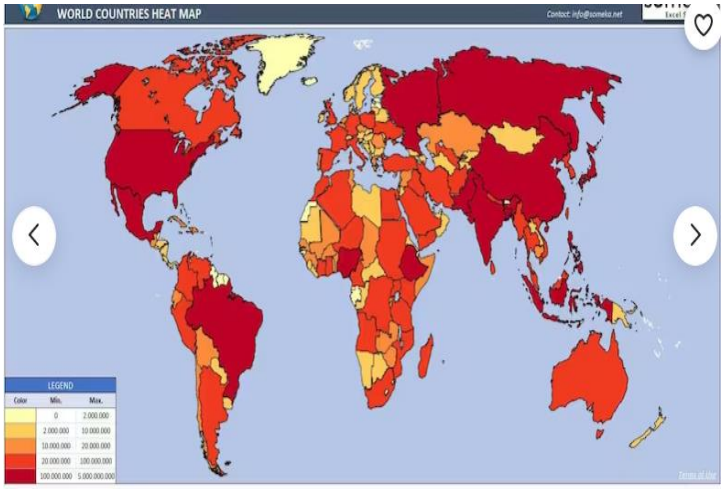

Figure 4: The Global Countries Heat Map

9.2 **Elements of the Earth Science Division:**

- Flight
- Analysis and Research
- Applied Earth Sciences
- Office of Earth Science and Technology
- Systems for ESD (Earth Science Data)

9.3 **Research and Analysis Program:**

On all spatial and temporal scales, our world is changing. The Earth Science Research and Analysis Program's objective is to broaden our understanding of Earth as a system and how it reacts to both natural and anthropogenic changes. Understanding the Earth as a complex system is necessary to comprehend the causes and implications of climate change and other significant global environmental concerns.

- **Atmospheric Elements:** The emphasis area on atmospheric composition studies the elements that make up Earth's atmosphere, notably the troposphere and stratospheric.
- **Atmospheric dynamics and Weather:** Events on a global scale can be predicted up to two weeks in advance in the weather system, which includes the focal region of the atmosphere and its interaction with oceans and land.
- **Cycle of Water and Energy:** Water and Energy Cycle researches how water and energy are transported, transformed, and distributed across the Earth System.

- **Variability and Change in the Climate:** This research provides global scale data from observation sets on Ocean and Ice.
- **Ecosystems and Carbon Cycle**: This research area focuses on Cyclic of carbons and ecosystems as it's changed naturally.
- **Earth's exterior and interior**: This research and attention area support the behaviors and features of the solid Earth, from the crust to the core.

- 10. **Some Powerful tools for Analytics in Earth Science:**
  - The Geographically Information System
  - Remote Sensing
  - GIS Techniques

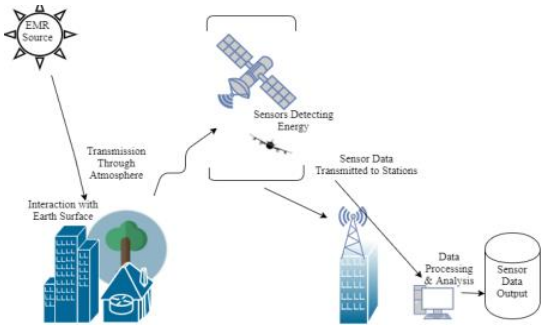

Figure 4: Remote Sensing

11. **The Challenge of Big Earth Data**: Earth is a complicated system that is always undergoing change. So, the challenge of Big Earth Data can be described by four vs.

- **Volume:** The datasets are substantial.
- **Variety:** datasets may include many data forms or kinds.
- **Velocity:** The data arrive quickly
- **Veracity:** The accuracy or availability of the data may be in question.

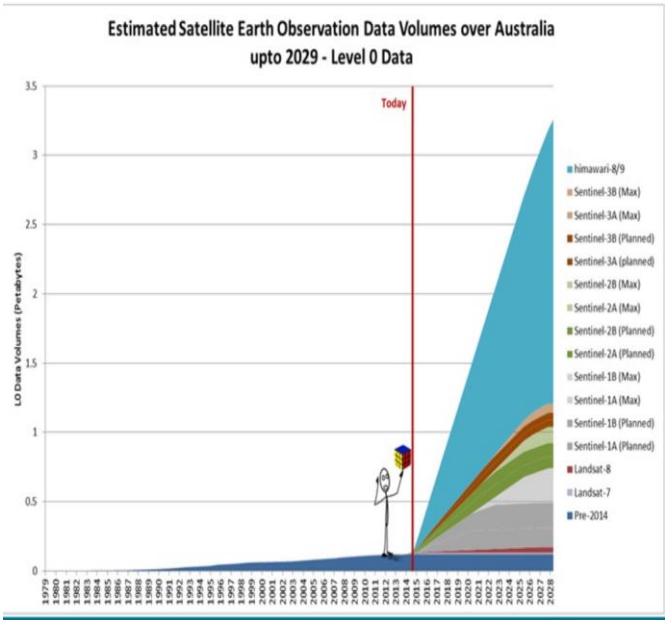

Figure 5: Satellite Earth Observation

12. **Conclusions**:

Earth science has developed into a variety of disciplines that incorporates aspect from not only the typical atmospheric component's but also related science like DS, ML, and DL in order to make sense of the huge volumes of data that are available in the field.

We have looked at Earth science in relation to the programs and tools used to look at Earth science data. As a result of higher resolution and the accessibility of unusual data sources, such as unstructured data, investigating and comprehending Earth science has taken on new dimensions as a result of technological advancements.

The techniques described here have been successfully used in research exploring and analysis of earth data.

13. REFERENCES

[1] Thomas Huang,Tiffany C.Vance," Big Data Analytics in Earth, Atmospheric, and Ocean Sciences" , January 2020

[2] Dr. Timothy Lang,"Observation Data" https://www.earthdata.nasa.gov/

[3] Tiffany C.Vance, Thomas Huang," Analyzing Big Earth data" https://eos.org/editors-vox/analyzing-big-earth-data-progress-challenges-opportunities

[4] Agrawal, D., Das, S., & El Abbadi, A. (2011, March). Big data and cloud computing: Current state and future opportunities

[5] Acker, J. G., & Leptoukh, G. (2007). Online analysis enhances use of NASA earth science data. Eos, Transactions

[6] Cressie, N. (2015). Statistics for spatial data. Hoboken, NJ: John Wiley & Sons

[7] Earth Programs https://science.nasa.gov/earth-science/programs/flight-programs

[8] Applications of Earth Programs (The European Space Agency) https://www.esa.int/Applications/Observing_the_Earth

[9] Research and Analysis Program (INDIAN SPACE RESEARCH ORGANISATION) https://www.isro.gov.in